MediaWiki file storage

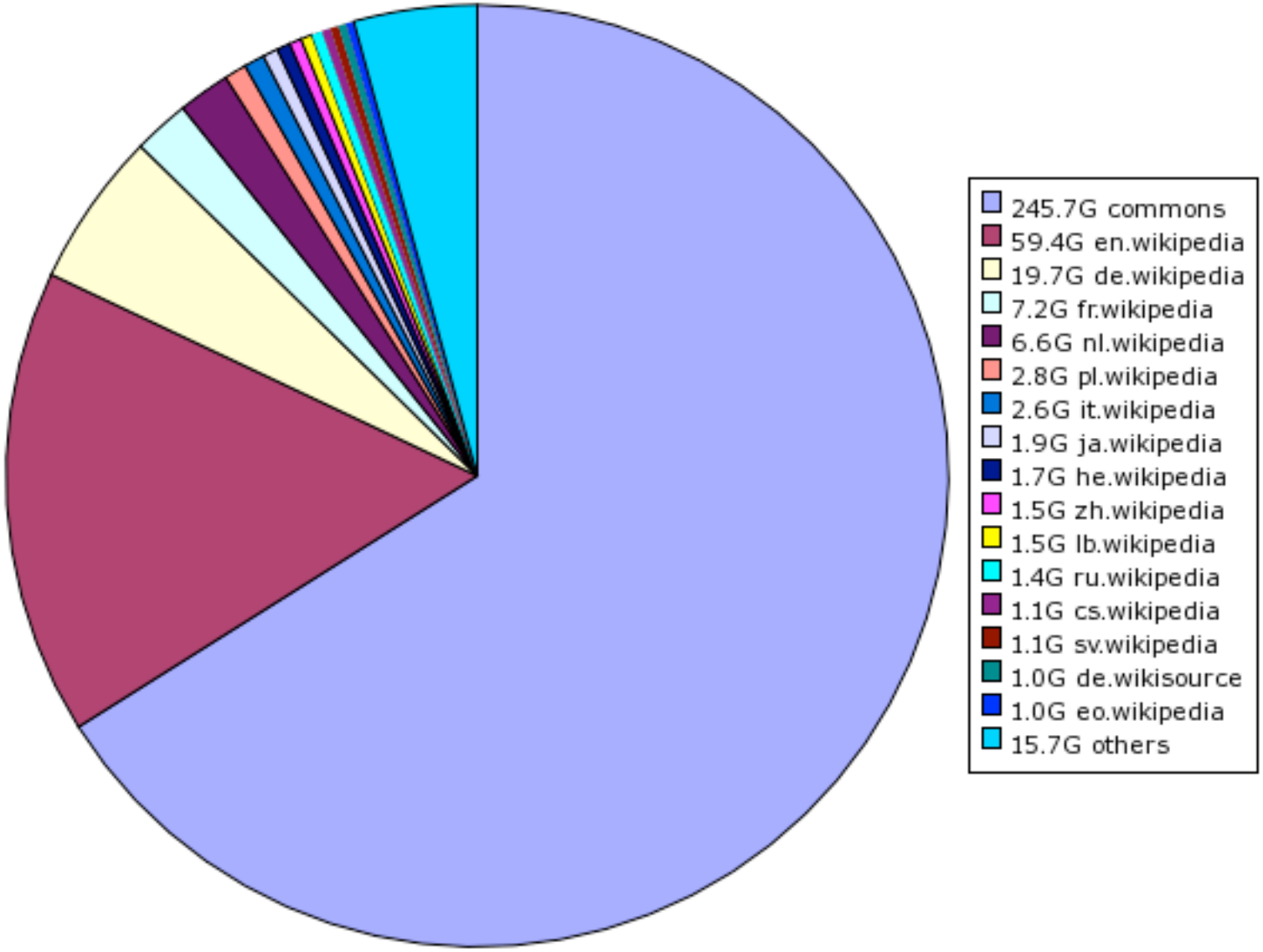
the files, they are a-changing

Brion Vibber
Wikimedia Foundation, Inc.

August 3, 2006

Wikimania Hacking Days
Cambridge, MA

372 gigs of uploaded files



- Before:
  - images/c/cf/Blarf_bloo_blee.jpg

    -or-

  - images/archive/c/cf/20060803022344!Awesome.jpg
- After:
  - images/0/7/d/07dn2l6f2u2qn5bd2mlvx7kvwjhoj1y2.jpg

Eww! Why's that good?

- URLs remain stable, simplifying caching
- Offsite copy-on-request mirrors will be possible
- Avoid storage waste for duplicate uploads
- Avoid storage waste for reverts
- Avoid fragile file copying on replacement upload
- Files could be renamed in the interface without worrying about the filesystem

Naming requirements

- base-16 of a 160-bit hash is 40 chars, very long
- base-64 smaller but not safe, if we want to work on case-insensitive filesystems
- Base-32 or base-36 fit in 32 characters
- One digit, three levels should work pretty well
- Hash subdirs should not include “ad” or other “words” that trigger common filters